Testing the validity of criminal risk assessment tools in sexually abusive youth

Barra, Steffen ; Bessler, Cornelia ; Landolt, Markus A ; Aebi, Marcel

Abstract: Although accurate risk appraisals are mandatory to provide effective treatment to juveniles who have sexually offended (JSOs), the current knowledge on the validity of risk assessment instruments for JSOs is inconclusive. We compared the predictive validities of the Juvenile Sex Offender Assessment Protocol II (J-SOAP II), the Estimate of Risk of Adolescent Sexual Offense Recidivism (ERASOR), and the Violence Risk Appraisal Guide-Revised (VRAG-R) scores concerning sexual, nonsexual-violent, and general criminal recidivism (based on both official and nonregistered reoffenses) in a consecutive sample of 597 male JSOs (Mage = 14.47 years, SDage = 1.57 years) while accounting for different recidivism periods, offense severities, and cumulative burden of adverse childhood experiences (ACEs). Receiver Operator Characteristic (ROC) curves and Cox regression analyses indicated that the tools allowed valid predictions of recidivism according to their intended purposes: The ERASOR was best suited to predict sexual recidivism within 0.5 and 3 years, the J-SOAP II was valid for predictions of sexual and nonsexual-violent recidivism within these recidivism periods, and the VRAG-R showed potential strengths in predicting nonsexual-violent recidivism, especially when committed above age 18. Elevated offense severity and burden of ACEs impeded predictive accuracies of the J-SOAP II and the VRAG-R, particularly in case of sexual recidivism. Our findings emphasize that risk assessment for JSOs must not rely solely on scores derived from risk assessment instruments, but a comprehensive consideration of a JSOs offense severity and psychosocial adversities is additionally necessary to approach accurate risk appraisals. (PsycINFO Database Record

DOI: https://doi.org/10.1037/pas0000590

Abstract

Although accurate risk appraisals are mandatory to provide effective treatment to juveniles who have sexually offended (JSOs), the current knowledge on the validity of risk assessment instruments for JSOs is inconclusive. We compared the predictive validities of the J-SOAP II, the ERASOR, and the VRAG-R scores concerning sexual, nonsexual-violent, and general criminal recidivism (based on both official and non-registered re-offenses) in a consecutive sample of 597 male JSOs (M_{age} = 14.47 years, SD_{age} = 1.57 years) while accounting for different recidivism periods, offense severities, and cumulative burden of adverse childhood experiences (ACE). Receiver Operator Characteristic (ROC) curves and Cox regression analyses indicated that the tools allowed valid predictions of recidivism according to their intended purposes: The ERASOR was best suited to predict sexual recidivism within 0.5 and 3 years, the J-SOAP II was valid for predictions of sexual and nonsexual-violent recidivism within these recidivism periods, and the VRAG-R showed potential strengths in predicting nonsexual-violent recidivism, especially when committed above age 18. Elevated offense severity and burden of ACEs impeded predictive accuracies of the J-SOAP II and the VRAG-R, particularly in case of sexual recidivism. Our findings emphasize that risk assessment for JSOs must not rely solely on scores derived from risk assessment instruments, but a comprehensive consideration of a JSO's offense severity and psychosocial adversities is additionally necessary to approach accurate risk appraisals.

Keywords: juvenile sexual offenders, J-SOAP II, ERASOR, VRAG-R, criminal recidivism

Public Significance Statement: The present study supports the use of the J-SOAP II, the ERASOR, and the VRAG-R to estimate the risk of criminal reoffending in sexually abusive adolescents. However, because risk prediction was impeded with greater offense severity and burden of childhood adversity, a comprehensive consideration of the adolescents' criminal and developmental histories is recommended over the exclusive reliance on risk assessment instruments in order to make conclusions about individual re-offense risks.

According to official statistics, almost every fifth sexual offense (apart from prostitution) registered in 2015 in the USA and Switzerland was committed by a minor (below age 18) (Federal Bureau of Investigation, 2016; Swiss Federal Statistical Office, 2016). Among juveniles who have been convicted or charged of sexual offenses (JSOs), recidivism rates range from 3-10% for sexual re-offenses up to 43-45% for non-sexual re-offenses (Aebi, Plattner, Steinhausen, & Bessler, 2011; Caldwell, 2016). Identifying those JSOs who are at highest risk of engaging in persistent crime is a main challenge for judicial and clinical professionals in order to ensure that JSOs receive appropriate interventions that prevent them from re-offending but avoid stigmatization (Miccio-Fonseca & Rasmussen, 2011; Parks & Bard, 2006; Prentky et al., 2010; Rasmussen, 2013). Whereas a number of instruments have been developed to estimate recidivism risk in adult offenders, only few risk assessment tools exist for application with JSOs, and research about their reliability and predictive validity is inconclusive (Hempel, Buck, Cima, & van Marle, 2013; Murphy, Page, & Hoberman, 2016).

Different approaches to assess risk of criminal recidivism

Three approaches are commonly discussed for the appraisal of risk of criminal recidivism in adolescent and adult offender samples: (a) unstructured clinical judgment (UCJ), (b) actuarial approaches, and (c) structured professional judgment (SPJ). UCJ is exclusively based on the assessor's clinical impression of the offender's individual recidivism risk, whereas actuarial approaches are structured methods of risk estimation that are characterized by a list of empirically-established (statistically weighted) risk and/or protective factors which contribute to overall sum scores (or risk categories) based on given algorithms. SPJ is also based on a range of empirically-established risk and protective factors but the overall recidivism risk is estimated by the assessor's individual weighting of these factors as well as his/her clinical impression based on the comprehensive consideration of an

individual's developmental, personal, and criminal characteristics (see Brown and Singh (2014) for more details about these three approaches and their specific advantages and disadvantages.

In their meta-analysis which included more than 45,000 adults and adolescents who had committed sexual offenses, Hanson and Morton-Bourgon (2009) found actuarial assessment instruments superior to UCJ concerning the prediction of sexual, violent, and general reoffending, whereas the predictive accuracies of SPJ procedures ranged between those of actuarial and UCJ approaches. However, the authors underscore that the number of studies that included SPJ was low and that the related findings were inconclusive, and thus suggest further examining the predictive validities of SPJ-based instruments (Hanson & Morton-Bourgon, 2009). Bearing in mind that SPJ is the most commonly used risk assessment approach for JSOs (Hempel et al., 2013), the investigation of risk assessment instruments and the comparison with different risk assessment approaches, such as actuarial approaches, is of major importance in samples of JSOs.

J-SOAP II and ERASOR

Two well-established JSO risk assessment tools that have received major scientific interest (Hempel et al., 2013; Miccio-Fonseca, 2016; Rettenberger, Klein, Martin, & Briken, 2014; Ryan, 2016) are the Juvenile Sex Offender Assessment Protocol II (J-SOAP II; Prentky & Righthand, 2003) and the Estimate of Risk of Adolescent Sexual Offense Recidivism (ERASOR; Worling & Curwen, 2001). The J-SOAP II (or, more exactly, its precursor J-SOAP (Prentky, Harris, Frizzell, & Righthand, 2000)) has been considered the "pioneer of risk assessment tools" for JSOs (Rasmussen, 2013, p. 124). First introduced in 1994 (Prentky et al., 2000), the instrument underwent a series of studies on its factor structure, reliability, and predictive validity, and was published in its current 28-item format in 2003 (Prentky & Righthand, 2003). The J-SOAP II is applicable to JSOs (with contact-offenses) between 12

and 18 years of age to predict sexual and nonsexual re-offending (Prentky & Righthand, 2003). As item scores are cumulated to create an overall risk score as well as specific subscale scores (see below), the J-SOAP II was developed in order to provide an actuarial risk assessment instrument for JSOs; however, it is currently not meant to be used as such because no cutoff values or risk categories have yet been established. Yet, since the J-SOAP II is based on risk factors for sexual and other criminal offending established in the literature, it may serve as a guidance tool for decision-making as part of an extensive SPJ risk assessment process (Prentky & Righthand, 2003). Due to potential changes in dynamic risk factors, re-assessment is recommended every six months or even more often when relevant changes occur in a JSO's living conditions (Prentky et al., 2010). Regarding the predictive validity of the J-SOAP II score interpretation, recent research points to mixed results (Fanniff & Letourneau, 2012; Wijetunga, Martinez, Rosenfeld, & Cruise, 2016). Although its developers highlight the predictive value of the scores of the J-SOAP II and its precursors for sexual reoffending (Prentky et al., 2010; Righthand et al., 2005), other researchers have reported more critical results related to the prediction of sexual and nonsexual recidivism (e.g., Caldwell, Ziemke, & Vitacco, 2008; Martinez, Rosenfeld, Cruise, & Martin, 2015; Viljoen et al., 2008; Wijetunga et al., 2016).

The ERASOR was developed based on the Sexual Violence Risk-20 scale (Boer, 1997) and empirically-proven risk factors for sexual offending (Worling & Curwen, 2001). According to the authors (Worling & Curwen, 2001), the ERASOR is applicable for risk assessment concerning sexual recidivism in JSOs (with contact-offenses) between 12 and 18 years of age. The authors highlight that the ERASOR is not suitable as an actuarial instrument but offers guidance for raters to make clinical decisions based on relevant risk factors. Thus, the ERASOR is a SPJ-based instrument. Re-assessment is recommended periodically (e.g., every six months; Viljoen, Elkovitch, Scalora, & Ullman, 2009) and is

necessary when relevant changes in risk factors occur. The authors themselves found the ERASOR scores to show good psychometric properties, the ERASOR clinical rating scores to be predictive of sexual recidivism within a period of 2.5 years, and a cumulative risk score (as the sum of all 25 items) to be predictive of sexual and non-sexual violent recidivism over a mean follow-up period of 3.66 years (Worling, 2004; Worling, Bookalam, & Litteljohn, 2012; Worling & Langton, 2015). However, other research groups found the ERASOR scores not to be predictive of sexual re-offending but suitable to predict general re-offenses (e.g., Quenzer & Dahle, 2010; Viljoen et al., 2009).

As to a direct comparison of the J-SOAP II and ERASOR scores, a meta-analysis of 33 studies including more than 6000 JSOs did not find any differences in their moderate predictive validities for sexual re-offending (Viljoen, Mordell, & Beneteau, 2012). In conclusion, both the J-SOAP II and the ERASOR may be used collaterally in risk assessment to approach a clinical decision but scores of neither of the tools have yet shown consistent empirically-proven reliability and validity (Miccio-Fonseca, 2016; Schlank, Matheny, & Schilling, 2016). Thus, more research on the usability of these instruments for risk assessment in JSOs is needed (Rasmussen, 2013).

VRAG-R

Although caution is advised in the comparison of risk factors for JSOs and adults who have sexually offended (Miccio-Fonseca & Rasmussen, 2011; Ryan, 2016; Schlank et al., 2016), recent research has indicated that the scores of risk assessment tools originally designed for adults may have some predictive values for JSOs too (Ralston & Epperson, 2013). Lately, the Violence Risk Appraisal Guide-Revised (VRAG-R; Harris, Rice, Quinsey, & Cormier, 2015; Rice, Harris, & Lang, 2013) has been introduced as a simplified, time-efficient revision of the Violence Risk Appraisal Guide (VRAG; e.g., Harris, Rice, & Quinsey, 1993) and the Sex Offender Risk Appraisal Guide (SORAG; e.g., Quinsey, Harris,

Rice, & Cormier, 1998) which represent actuarial tools to assess the risk of violent (including sexual) reoffending intended to use for adolescent and adult sexual and violent offenders. The VRAG and SORAG scores had been found to predict violent and sexual re-offenses in adult samples, especially for older offenders (above age 25) and those with sexual offenses of greater severity (Rettenberger & Eher, 2007; Singh, Grann, & Fazel, 2011). Still, although its authors recommended to replace the VRAG and SORAG by the VRAG-R (Harris et al., 2015), studies on the accuracy of the VRAG-R scores are scarce. The authors themselves pointed to satisfactory predictive validity of the VRAG-R scores in adult and adolescent samples (Harris et al., 2015; Rice et al., 2013). However, in comparison, a recent independent investigation of the VRAG-R's performance in 120 adult offenders who had been released from a correctional facility in Canada revealed somewhat smaller values of an overall moderate predictive validity for the VRAG-R scores concerning violent and general reoffending (Glover, Churcher, Gray, Mills, & Nicholson, 2017). In a sample of adult forensic inpatients from Canada, scores of the VRAG-R were not found to predict inpatient aggression (Hogan & Olver, 2016).

Although the performance of the VRAG-R had been tested in adult sex offenders and general offenders younger than 18 years and the SORAG had been tested in both sexually abusive adults and adolescents (Harris et al., 2015), to the best of our knowledge, no studies have been published to date on the predictive validity of the VRAG-R scores for JSOs. Taking into account the proposed applicability of the VRAG-R for both sexual and non-sexual violent offenders to predict violent (including sexual) reoffending as well as the rarity of actuarial risk assessment tools for JSOs, it appears beneficial to examine whether the VRAG-R may qualify for a valid risk assessment concerning sexual and other violent recidivism in JSOs. So far, the developers of the VRAG-R have suggested using the VRAG-R in adolescent offender samples only for the assessment of violent (including sexual)

recidivism in adulthood (above age 18), although research to support this recommendation is still lacking. In conclusion, we decided to include three risk assessment instruments – two of them well-established and one newly-constructed - in order to both replicate but also advance previous research on risk assessment in JSOs.

Considering offense severity and adverse childhood experiences in risk prediction

Some researchers have criticized applying risk assessment instruments to samples that differed from those the tools were initially constructed for, e.g., in regard to offense severity. Hecker (2014), for instance, underscored that the J-SOAP II had been developed in a sample of JSOs who had shown coercive sexual (contact) activity, and thus questioned its applicability in JSOs with somewhat less severe (e.g., non-contact) offenses. Although some studies have compared the predictive validities of risk assessment scores in JSO samples with higher proportions of severe offenses (e.g., JSOs in correctional facilities) to the predictive validities of these instruments' scores in JSO samples with higher proportions of less severe offenses (e.g., JSOs in residential treatment programs), findings are inconclusive (e.g., Martinez et al., 2015). Moreover, we are not aware of any study that has directly examined the potential effects of offense severity on risk prediction in JSOs. Bearing in mind that greater offense severity has been found to play a non-negligible role in the prediction of criminal recidivism in JSOs (e.g., Aebi et al., 2011), more research is needed to examine the accuracy of risk assessment tools in consecutive samples with varying degrees of offense severity.

Furthermore, recent empirical findings have highlighted that recidivism in general juvenile delinquents and in JSOs may be associated with the adolescents' histories of adverse childhood experiences (ACEs) (e.g., Baglivio, Wolff, Piquero, & Epps, 2015; Carpentier & Proulx, 2011; Mallie, Viljoen, Mordell, Spice, & Roesch, 2011; van der Put & De Ruiter, 2016). Fox, Perez, Cass, Baglivio, and Epps (2015) even point to the potential use of a

cumulative ACE score (sum of multiple experienced ACE categories) in the early detection of future severe, violent, and permanent adolescent offenders. Other researchers have also recommended considering the influence of ACEs (e.g., in terms of the adolescents' "parents' issues, denial, support, or lack of support"; Schlank et al., 2016, p. 255) in juvenile risk assessment. As most JSOs have been found to be burdened with multiple ACEs (e.g., Barra, Bessler, Landolt, & Aebi, 2017) and to show more types of ACEs than general adolescent offenders (Seto & Lalumière, 2010), the consideration of ACEs in risk assessment appears to be of particular importance for these juveniles. Although risk assessment tools for JSOs usually include some types of ACEs (e.g., J-SOAP II: item 8 - *sexual victimization*, item 16 - *physical victimization/family violence*, item 23 - *quality of peer relationships*, item 28 - *support system*; ERASOR: item 13 - *social isolation*, item 20 - *family dysfunction*, item 21 - *parental refusal*; VRAG-R: item 1 - *living with both biological parents*), none of them provides a comprehensive consideration of a JSO's ACE history. Thus, it remains unclear whether cumulated intra- and extra-familial ACEs would influence their scores' predictive validity for criminal recidivism.

Limitations of previous research

So far, empirical evidence on the predictive validity of the scores of risk assessment tools for JSOs is mixed (e.g., Hempel et al., 2013). This might be partly due to the fact that sample sizes are usually quite small and the base rates of sexual recidivism considerably low (Aebi et al., 2011; Fanniff & Letourneau, 2012; Hempel et al., 2013; Miccio-Fonseca, 2016; Parks & Bard, 2006; Wijetunga et al., 2016; Worling et al., 2012). Furthermore, the rapid developmental changes in adolescents' risk factors make it difficult to derive long-term risk predictions, and thus call for the inclusion of different recidivism periods when investigating the predictive accuracies of risk assessment tools (Fanniff & Letourneau, 2012; Hempel et al., 2013; Miccio-Fonseca, 2016; Ralston & Epperson, 2013; Schlank et al., 2016; Viljoen et al.,

2012). It is also worth noticing that, if risk assessment is conducted for the purpose of intervention planning in the course of initial court procedures, validation studies will have to rely on samples of JSOs who have not yet received any measure. This is rarely the case, as most researchers have examined JSOs from treatment conditions (e.g., Martinez et al., 2015; Rajlic & Gretton, 2010; Viljoen et al., 2009; Viljoen et al., 2008; Worling et al., 2012), correctional facilities (e.g., Caldwell et al., 2008; Martinez et al., 2015; Parks & Bard, 2006) , or at the time of release from a judicial measure (e.g., Wijetunga et al., 2016). Only few studies have included consecutive samples or JSOs on probation (e.g., Aebi et al., 2011; Ralston & Epperson, 2013; van der Put, van Vugt, Stams, Deković, & van der Laan, 2013; Worling & Langton, 2015). It is also important to note that a considerable number of adolescents may not have conducted contact offenses but non-contact sexual harassment. In fact, non-contact (e.g., Internet-based) offenses have been found the most prevalent type of sexual offending among juveniles (e.g., Mohler-Kuo et al., 2014). Taking into account that non-contact offenses may as well lead to legal charges and/or convictions, it appears beneficial to (a) include JSOs with non-contact offenses in the investigation of the predictive validities of the scores of risk assessment tools; and (b) include non-contact sexual offending in measures of sexual recidivism. To the best of our knowledge, no risk assessment instruments are yet available which specifically consider non-contact offending in JSOs. All risk assessment instruments examined in this study have been originally developed in samples of contact offenders. We therefore aimed to investigate a wider and more representative sample of JSOs and expand previous research on risk assessment for JSOs by including measures of non-contact offenses. Moreover, we are not aware of any studies that have examined the direct and interactive effects of offense severity and/or the accumulation of adverse childhood experiences on the predictive accuracies of the scores of risk assessment instruments for JSOs.

The present study

Addressing the shortcomings of previous research mentioned above, the present study is the first to test the predictive accuracy of the J-SOAP II, the ERASOR, and the VRAG-R simultaneously in a large consecutive sample of JSOs while considering different recidivism intervals, the degree of offense severity, and the cumulative burden of ACEs.

Based on previous research and the purposes described by its authors, we expected scores of the J-SOAP II and the ERASOR to show comparable predictive validities for sexual recidivism and scores of the J-SOAP II to be more accurate in the prediction of nonsexual recidivism. Bearing in mind potential developmental changes during adolescence, we expected predictions to be more accurate with greater temporal proximity. The VRAG-R scores were assumed to predict sexual and nonsexual violent re-offending, however, with more accuracy for re-offenses committed in adulthood (above age 18) than in adolescence. As the risk assessment tools were constructed on samples with rather high degrees of offense severity, we expected their scores' predictive accuracy to be higher in JSOs with more severe offenses. Due to the limited empirical foundation, analyses concerning the effects of a cumulative ACE score on risk prediction were performed in exploratory manner.

Methods

Measures

J-SOAP II. The J-SOAP II (Prentky & Righthand, 2003) contains 28 items that reflect risk factors assigned to four subscales: (1) Sexual Drive/Preoccupation (items 1-8), (2) Impulsive/Antisocial Behavior (items 9-16), (3) Intervention (items 17-23), and (4) Community Stability/Adjustment (items 24-28). The first two and the last two subscales can be combined into two higher order scales displaying static and dynamic risk factors, respectively. Moreover, a J-SOAP total score is calculated by adding the risk estimates of all 28 items. Each item is to be rated between 0 and 2 with higher scores representing increased

severity. Rating instructions are provided in the coding manual. Items were rated against lower risk when pertinent information from the court files was insufficient. For reasons of comparability with the other instruments, we only included the J-SOAP total score (ICC = .74) in our current analyses. Since the total score is built by a summation of the ratings on the given risk factors, the J-SOAP II was used as a mechanical instrument. We used the German translation of the J-SOAP II (Schmelzle, 2004) in the present study.

ERASOR. The ERASOR (version 2.0; Worling & Curwen, 2001) contains 25 items from the domains (1) Sexual Interests, Attitudes, and Behaviors (items 1-4), (2) Historical Sexual Assaults (items 5-13), (3) Psychosocial Functioning (items 14-19), (4) Family/Environmental Functioning (items 20-23), and (5) Treatment (items 26-28). However, these domains are rather descriptive in purpose and do not represent specific subscales. There is additional space to include one further risk factor which may be of specific importance for an individual JSO. Each item is to be rated as present, partially/possibly present, not present, or unknown. Rating instructions are provided in the coding manual. Items were rated against lower risk when pertinent information from the court files was insufficient. The overall risk is to be estimated in form of a clinical judgment (based on a personal weighting of given risk factors) as low (= 0), moderate (= 1), or high (= 2) (ICC = .60). We used the German translation of the ERASOR (Schmelzle, 2003) in the present study.

VRAG-R. The VRAG-R (Harris et al., 2015; Rice et al., 2013) contains 12 items that cover different risk factors for sexual/violent re-offending such as elementary school maladjustment, history of alcohol and drug problems, criminal history, or conduct disorder. Rating instructions are provided in the coding manual. Each item is rated with specifically developed weighting with an overall range from -6 (item 12) to +6 (items 9 and 12), in which higher scores represent increased severity. Items were rated against lower risk when pertinent information from the court files was insufficient. When raters omitted to score an item, a

prorating procedure was applied as proposed by the VRAG-R's authors (Harris et al., 2015). The sum of item ratings represents the VRAG-R total score (ICC = .92) which was used in the present study. We did not use cutoff-based risk categories for the sake of comparability with the other instruments, and because such risk categories have not yet been validated in adolescent samples.

In the adult version of the VRAG-R, antisociality (item 12) is represented by Facet 4 of the Psychopathy Checklist-Revised (Hare, 2003). However, as suggested by the authors (Harris et al., 2015), we used Facet 4 of the Psychopathy Checklist Youth Version (Forth, Kosson, & Hare, 2003) instead for our adolescent sample. Because no translation was yet accessible for the VRAG-R at the time data were assessed, we used the original version in the present study (recently, Rettenberger, Gregório Hertz, and Eher (2017) have published a German translation of the VRAG-R).

Offense severity. We assessed offense severity by means of a scale by Aylwin et al. (2000), which was originally developed to describe differences in offense characteristics between JSOs and adult sexual offenders. An offender's most serious sexual assault is ranked according to specific offense characteristics independent of their effects on the victim(s). The original coding contains six levels with increasing severity: (1) obscene phone calls, voyeurism, and clothed fondling, (2) exhibitionism, frotteurism, and clothes-off fondling, (3) oral and/or simulated sex, (4) attempt/performance of vaginal sex, (5) attempt/performance of anal sex or gang rape, and (6) use of augmented (physical and psychological) force. In order to differentiate between non-contact and contact assaults, we divided the first category into two sections reflecting non-contact (e.g., obscene phone calls, sexual harassment via the Internet) and contact (e.g., clothed fondling) assaults (Aebi et al., 2011). Thus, the final scale used in the present study contained seven levels indicating elevating degrees of offense severity (ICC = .79). Offense severity was rated according to the descriptions of a JSO's

index offense given in his case files. When a JSO had been convicted for multiple sexual offenses, the most serious offense was rated. To the best of our knowledge, there is no validation study to date for this severity scale. However, the scale has been successfully applied in previous research on JSOs (Aebi, Plattner, Steinhausen, & Bessler, 2011).

Adverse childhood experiences. Adverse childhood experiences (ACEs) that had occurred before a JSO's first currently convicted sexual assault were assessed following the ten categories of the Maltreatment and Abuse Chronology of Exposure (MACE) scale (Isele et al., 2014; Teicher & Parigger, 2015): verbal abuse, non-verbal emotional abuse, physical abuse, emotional abuse by peers, physical bullying, emotional neglect, physical neglect, witnessing violence between caretakers, witnessing violence against siblings, and sexual victimization. We further relied on the Child Sexual Abuse Questionnaire (CSAQ; Mohler-Kuo et al., 2014) to amplify information on histories of sexual victimization by including items related to a range of non-contact (online) assaults (because of the frequent appearance of non-contact (online) assaults among adolescents; Mohler-Kuo et al., 2014). An ACE-category was assured when any pertinent experience was recorded in the court files. Assured ACE-categories were counted to create a total ACE score (potential range: 0 – 10; ICC = .86). MACE and CSAQ scores were proven to be reliable and valid for the assessment of ACEs in self-reports (Aebi et al., 2015; Isele et al., 2014; Teicher & Parigger, 2015). The present inter-rater agreement supports their use in file-analyses.

Criminal recidivism. Two sources were used to derive information on criminal recidivism. First, we included official re-offenses registered by the Swiss Federal Office of Justice and the Swiss Federal Statistical Office. Second, we additionally assessed any delinquent behavior (meeting the criteria to be potentially charged under Swiss penal law) which a JSO had shown after the current conviction and which was recorded in the case files but may not have led to new charges and/or convictions (e.g., verbal or physical sexual

harassment or other physical violence against or stealing things from peers or staff in an institution that a JSO had lived in). By including such information by professionals working with the JSOs, e.g. from probation reports or therapeutic documentations, we aimed to approximate the actual crime prevalence in the dark field (e.g., Thornberry & Krohn, 2000). To counteract potential bias, re-offense data were assessed after coding all other variables of interests. Three categories of criminal recidivism were considered: (1) sexual recidivism, i.e., any sexually offending behavior except illegal pornography use[1] ($\kappa_{official}$ = .89; $\kappa_{case\ files}$ = .79; (2) nonsexual-violent recidivism, e.g., violence-related threat, bodily assault, affray, robbery ($\kappa_{official}$ = 1.00; $\kappa_{case\ files}$ = .60); and (3) general recidivism, including the first two recidivism categories as well as further contraventions such as drug-related offenses, theft, or property damage ($\kappa_{official}$ = 1.00; $\kappa_{case\ files}$ = .80). In order to examine short- and long-term predictions of risk assessment tools, recidivism was examined 0.5 years (183 days) and 3 years (1095 days) after the current conviction. The recidivism period of 0.5 years is in agreement with the proposed time interval for re-assessing re-offense risk with the J-SOAP II and the ERASOR (Prentky et al., 2010; Viljoen et al., 2009). The maximum of 3 years was chosen because recent research has indicated that re-offense rates tended to stagnate at this time (Caldwell, 2016). We additionally considered re-offenses exclusively conducted in adulthood (above age 18) to test the recommendation to use the VRAG-R for adolescent risk assessment related to adult recidivism (Harris et al., 2015).

Procedure

Between February and December 2015, court files of all adolescents were analyzed who had been convicted for a sexual offense apart from pornography (e.g., rape, sexual coercion, sexual molestation of a child, exhibitionism, sexual harassment)[1] between January 2007 and September 2014 in 14 German-speaking cantons (states) in Switzerland. A forensic psychologist, a PhD candidate in forensic psychology, and a psychology student at the

master's degree level conducted data assessment using a specifically developed, structured coding manual based on the Forensic Psychiatric Documentation System (Nedopil, Grassl, & Mende, 1986). In addition to biographical and offense-related information, the coding manual contained the J-SOAP II, the ERASOR, the VRAG-R, a section to assess ACEs, and a separate chapter to record on sexual and non-sexual re-offenses. Thirty court files were selected randomly under consideration of file content (inclusion of psychiatric/psychological expert opinion) and residential area (more or less than 10,000 residents) to be blindly double-rated by the two forensic psychologists in order to calculate inter-rater agreement (Cohen's κ for nominal and the intra-class correlation coefficient [ICC; two-way random model, single measure, absolute agreement] for metric variables). A threshold of .60 was defined for κ and ICC to indicate substantial agreement (Fleiss, 1981; Landis & Koch, 1977)[2]. Study procedures were authorized by the ethics committees of Zurich and northwest/central Switzerland (lead ethics committee: Zurich, EC-No. 2010-0483) as well as all juvenile justice authorities involved.

Participants

Court files of a total of 687 JSOs were coded. Because the J-SOAP II and the VRAG-R were constructed for risk assessment in male offenders, female JSOs were excluded (n = 14, 2.0%). Male JSOs who were younger than 12 years at the time of their first currently convicted sexual assault (n = 75, 11.1%) were excluded too because the J-SOAP II and the ERASOR were constructed for risk assessment in JSOs between 12 and 18 years old. In addition, one male JSO had to be excluded due to data loss. The final sample for the present study contained 597 male JSOs between 12 and 18 years at the time of their first currently convicted sexual assault (M = 14.47, SD = 1.57). Convicted sexual assaults included verbal/online forms of sexual harassment (e.g., sexual explicit chatting, taking/distributing sexual explicit photos/videos without permission; n = 150, 25.1%), sexual molestation in a

face-to-face situation (e.g., exposing genitals, compelling victim to undress/present genitals, compelling victim to look at his or others' sexual activities; n = 324, 54.3%), nonconsensual touching (JSO touched the victim with sexual intent or had the victim touch him; n = 421, 70.5%), oral penetration (JSO conducted any oral sexual activity on the victim or had the victim conduct any oral sexual activity on him; n = 154, 25.8%), and/or vaginal/anal penetration (n = 125, 20.9%). Out of the total sample of JSO, 193 (32.3%) committed one or multiple sexual non-contact offenses only, whereas 404 (67.7%) committed at least one sexual contact offense.

Excluded male JSOs younger than 12 years differed from included JSOs in the proportions of offenses including verbal/online harassment (13.3%), $\chi^2(1) = 5.11$, $p = .024$, offenses including face-to-face molestation (70.7%), $\chi^2(1) = 7.27$, $p = .007$, offenses including oral penetration (48.0%), $\chi^2(1) = 16.20$, $p \leq .001$, and offenses including vaginal/anal penetration (42.7%), $\chi^2(1) = 17.57$, $p \leq .001$. No differences were found in rates of offenses including nonconsensual touching (70.7%), $\chi^2(1) = 0.01$, $p = .979$.

Statistical Analysis

Analyses were conducted in IBM SPSS 23. Parametric and non-parametric descriptive statistics included χ^2-tests, Pearson-, and Spearman-Rank correlations. The general level of significance was set at $p \leq .05$. Z-transformed scores were used for comparative analyses to account for scaling differences and to facilitate interpretation.

ROC analyses. The predictive accuracy of risk assessment scores was examined using Receiver Operator Characteristic (ROC) curves (e.g., Mossman, 1994). ROC curves describe the rate of true positives (sensitivity) in relation to the rate of false positives (1-specificity) for each possible score on an instrument's scale. The area under the curve (AUC) represents an indicator of overall predictive accuracy of the scores of the instrument, which allows comparison with other risk assessment tools. AUCs reflect the probability that the risk

score of a randomly chosen reoffender would be higher than the risk score of a randomly chosen non-reoffender. AUCs have been recommended as "preferred measure of predictive or diagnostic accuracy in forensic psychology and psychiatry" (Rice & Harris, 2005, p. 618), especially for risk assessment tools without cutoff-values (Singh, 2013). AUCs are base-rate independent and irrespective of tendencies against or toward type 1 and 2 errors in clinical evaluations (Mossman, 1994). We followed recommendations by Rice and Harris (2005) to interpret AUC values analogous to Cohen's (Cohen, 1988) effect size *d*: small (AUC = .556-.639; d = .20-.50), moderate (AUC = .639-.714; d = .50-.80), and large (AUC ≥ .714; $d \geq$ .80). Hanley and McNeil's (1983) non-parametric approach was used the compare the instruments' AUC values. Z-values above 1.96 or below -1.96 represented significant differences.

Cox regression. Predictive validity was further examined using Cox-regression. For each recidivism category (sexual, nonsexual-violent, general), time at risk (number of days between the current conviction and the first re-offense) was included as dependent variable. When a JSO did not re-offend, his time at risk was set at the maximum of 1095 days for the 3-year observation period. The assumption of proportional hazard (time-independent influence of covariates on individual hazard) was considered as confirmed when *p*-values exceeded the threshold of .05. Multicollinearity was assumed with variance inflation factors (VIFs) above 10 and tolerance scores below .10 (Hair, Anderson, Tatham, & Black, 1995).

Results

Descriptive Findings

Risk assessment tools. In the total sample (N = 597), the J-SOAP II total score (Cronbach's α = .91) ranged between 0 and 48 (M = 14.91, SD = 10.24). Based on the ERASOR overall risk score, raters assigned 260 JSOs (43.6%) to low, 242 (40.5%) to moderate, and 95 (15.9%) to high recidivism risk. The VRAG-R score (Cronbach's α = .73)

ranged from -23 to 34 (*M* = -10.16, *SD* = 11.13). Offense severity levels ranged from 1 to 7 (*Mdn* = 3), and the ACE score ranged from 0 to 10 (*M* = 1.74, *SD* = 2.17). Correlations among scores are presented in Table 1.

Recidivism rates. Rates of sexual, nonsexual-violent, and general recidivism within 0.5 years, 3 years, and above age 18 are shown in Table 2. All JSOs were followed for 0.5 years, 514 (86.1%) for 3 years, and 550 (92.1%) until adulthood (above age 18). Compared with reliance on official re-offenses only, the consideration of case file information which included non-registered re-offenses led to increased prevalence rates of (a) sexual recidivism within all observation periods, (b) non-sexual violent recidivism within 0.5 and 3 years, and (c) general recidivism within 0.5 years.

Predictive accuracy

Results of the ROC analyses for sexual, nonsexual-violent, and general recidivism at 0.5 and 3 years after the current conviction as well as above age 18 are displayed in Table 3[3]. J-SOAP II and ERASOR scores showed large predictive values for juvenile sexual recidivism in both the 0.5 and the 3 year observation periods, whereas the predictive value of the VRAG-R scores fell in the moderate range. Scores of all instruments showed moderate predictive values for sexual recidivism above age 18. Subsequent analyses revealed that the ERASOR scores predicted sexual recidivism within 3 years significantly better than the VRAG-R scores ($z = 2.15$).

Nonsexual-violent recidivism was predicted well by the J-SOAP II and the VRAG-R scores throughout all recidivism periods. The ERASOR scores showed moderate predictive values for nonsexual-violent recidivism. Subsequent analyses revealed that the J-SOAP II scores predicted nonsexual-violent recidivism consistently better than the ERASOR scores (0.5 years: $z = 2.28$; 3 years: $z = 2.92$; above age 18: $z = 2.20$). The VRAG-R scores

performed better than the ERASOR scores concerning nonsexual-violent recidivism above age 18 ($z = 2.50$).

General recidivism within 0.5 years was predicted well by the J-SOAP II and the VRAG-R scores. Yet, the VRAG-R scores showed moderate predictive validities for recidivism within 3 years and above age 18, whereas the J-SOAP II scores showed moderate predictive validity for recidivism within 3 years and small predictive validity for recidivism above age 18. The ERASOR scores predicted general recidivism with moderate accuracies for the 0.5 and 3 year observation periods but with small accuracy for re-offenses committed above age 18. Subsequent analyses revealed that the J-SOAP II scores predicted general recidivism within 3 years significantly better than the ERASOR scores ($z = 2.07$).

Effects of offense severity and cumulated ACEs

Tables 4-6 display the results of univariate and multivariate Cox regressions for the J-SOAP II, ERASOR, and VRAG-R, respectively. The proportional hazard assumption was confirmed for all variables (all $p > .05$). No multicollinearity issues emerged. Regression models supported the findings from previous ROC analyses: Risk scores on each instrument were positively related to (time to first) reoffending in all models. Concerning the prediction of sexual recidivism, (marginally) significant negative interactions between risk scores and offense severity were found for the J-SOAP II and the VRAG-R but not for the ERASOR scores. Significant negative interactions between risk scores and the cumulated ACE score were found for VRAG-R concerning sexual recidivism and for the J-SOAP II concerning sexual and general recidivism. When both interactions were concertedly considered in one regression model (model 5), the interaction effects of risk scores and ACE scores remained significant, and ACE scores usually maintained their main effects on recidivism risk.

Discussion

The current study is the first comparison of the predictive validities of the scores of the J-SOAP II, ERASOR, and VRAG-R in a consecutive sample of JSOs accounting for different recidivism periods, offense severities, and histories of ACEs. Overall, the present risk assessment tools proved to be valid for the prediction of criminal recidivism according to their intended purposes in a heterogeneous sample of JSOs. Yet, predictive validity was confined by elevated offense severity and, especially, by increasing numbers of ACEs.

Recidivism rates in the present sample

Recidivism rates for sexual, nonsexual-violent, and general recidivism were comparable to those found in previous Swiss and international samples (e.g., Aebi et al., 2011b; Caldwell, 2016) and indicated that JSOs more often re-offend with nonsexual criminal conduct than sexual assaults. The inclusion of non-registered reoffending contributed to elevated rates of recidivism with temporal proximity to the current conviction, especially in terms of sexual reoffending[4]. Recidivism studies should therefore not rely only on officially recorded re-offenses but include other sources of information (e.g., from case files or direct self-reports) to approximate the actual recidivism rates in the dark field (e.g., Thornberry & Krohn, 2000).

Predictive validities of risk assessment scores

J-SOAP II and ERASOR. Overall, the performances of the J-SOAP II and the ERASOR in the present study appeared to be better than those reported in most previous studies (e.g., Aebi et al., 2011b; Martinez et al., 2015; Quenzer & Dahle, 2010; Viljoen et al., 2009; Viljoen et al., 2012; Viljoen et al., 2008; Worling & Langton, 2015). This finding may be attributable to two facts: First, the present sample was more heterogeneous in terms of offense characteristics and recidivism risk than previous JSO samples. Increased sample homogeneity (e.g., due to pre-selection by offense characteristics or treatment setting) can

cause reduction in AUC estimates (Howard, 2016). Second, the consideration of both registered and non-registered re-offenses may have contributed to higher accuracy ratings. In fact, examining the relations of J-SOAP II scores and non-registered sexual recidivism in a sample of adolescents involved in the child welfare system, Prentky et al. (2010) found AUC values of up to .83.

In line with our expectations, the J-SOAP II scores proved to be valid in predicting sexual and nonsexual recidivism 0.5 to 3 years after the current conviction, whereas the ERASOR scores showed comparable (slightly higher) accuracy values for the prediction of sexual recidivism but lower accuracy values for nonsexual recidivism. Indeed, the J-SOAP II significantly outperformed the ERASOR in the prediction of general reoffending within a 3-year recidivism period. For each single assessment tool, predictive accuracies did not notably vary between 0.5-year- and 3-year recidivism periods, indicating that both the J-SOAP II and the ERASOR may be used for short- and long-term predictions. However, when only considering those re-offenses that JSOs had committed above age 18, we found that the predictive validity of the J-SOAP II scores was somewhat weaker, although the effect for nonsexual-violent recidivism was still large. In contrast, the ERASOR scores only showed low to moderate predictive effects for any category of recidivism above age 18, underscoring that its use for the prediction of reoffending in adulthood remains questionable.

For the interpretation of our results, one needs to bear in mind that we could not address changes in risk factors potentially influencing recidivism risk that might have occurred after risk assessment was completed. We therefore recommend not diverging from previous researchers' advice to periodically re-assess recidivism risk (Prentky et al., 2010; Viljoen et al., 2009).

VRAG-R. In the present first examination of the VRAG-R in a JSO sample, the instrument's scores proved to be more accurate in predicting nonsexual recidivism than

sexual recidivism. Predictive effects were similar to those found in previous adult sex offender and general juvenile offender samples relating to violent reoffending (Harris et al., 2015; Rice et al., 2013). As to a direct comparison, the ERASOR significantly outperformed the VRAG-R in the prediction of sexual recidivism within 3 years. Yet, the VRAG-R significantly outperformed the ERASOR in the prediction of nonsexual-violent recidivism in adulthood. Although its effects were slightly higher for sexual and nonsexual-violent recidivism exclusively committed above age 18, the moderate to large predictive effects of the VRAG-R scores for sexual recidivism and nonsexual-violent/general recidivism within the 0.5 and 3-year recidivism periods argue in favor of its use in juvenile offender samples not only for the prediction of adult recidivism but also for predicting more proximate reoffending.

The influence of offense severity and adverse childhood experiences on risk prediction

In case of the J-SOAP II and the VRAG-R, predictive accuracy for sexual recidivism appeared impeded when offense severity increased. These findings were unexpected in view of the fact that both instruments were developed in samples showing rather severe criminal conduct. However, elevated offense severity in JSOs has been associated with a range of disadvantageous factors (e.g., behavioral problems and prior delinquency; Leroux, Pullman, Motayne, & Seto, 2016), which may contribute to increased risk of re-offending but may not be adequately covered by the J-SOAP II and the VRAG-R. In contrast, the ERASOR may not have been influenced by offense severity as much as the other instruments because of its SPJ-approach in which the clinical impression underlying the final rating may have implicitly included such factors.

In the present sample, increased offense severity was associated with elevated rates of ACEs. This finding is in line with previous research that found multiply burdened JSOs to commit more severe offenses than JSOs with little or no ACE occurrence (e.g., Barra et al.,

2017; Burton, 2003). At the same time, the number of ACEs was associated with elevated recidivism risk, underscoring the role of ACEs in the incidence of criminal persistence in JSOs (e.g., Carpentier & Proulx, 2011; Mallie et al., 2011). Moreover, our findings emphasize that risk assessment is particularly challenging in JSOs who are highly burdened with ACEs: Predictive validities of the J-SOAP II and the VRAG-R concerning sexual recidivism decreased with increasing rates of ACEs, even when the interaction with offense severity was accounted for. The present findings suggest that a more comprehensive consideration of the cumulative effects of intra- and extra-familial ACEs may be beneficial. Further revisions of the J-SOAP-II may consider more detailed ACE profiles in order to increase the validity in highly burdened youth.

Actuarial vs. SPJ-based approaches

In total, the present findings underscore that a comprehensive consideration of a JSO's personal and criminal history (e.g., in terms of SPJ-based approaches) is of particular importance for risk prediction in JSOs. The mechanical (actuarial) summation of empirically-derived risk factors to predict criminal recidivism in case of the J-SOAP II and the VRAG-R was challenged when other risk factors, especially ACEs, were taken into account. In contrast, the SPJ-based assessment of the ERASOR was not significantly worsened with increasing offense severity and/or increasing numbers of ACEs. However, ACEs themselves were represented by a mechanical sum score in the present study, raising the question of whether the inclusion of such an ACE score may actually add incremental validity to risk assessment tools. Although our data support this suggestion, further research needs to examine whether it holds across different samples, different numbers and/or types of ACEs, and different risk assessment tools. However, bearing in mind the lack of reliability and validity of UCJ approaches (Brown & Singh, 2014), we recommend using existing risk

assessment instruments for guidance toward a clinical estimation of risk (in terms of SPJ), but relying neither on their isolated scores nor one's own clinical impression alone.

Strengths and limitations

The present study overcame some of the limitations noted in previous studies on the predictive validity of risk assessment instruments in JSOs. First, we examined a large consecutive, court-based sample of JSOs. Thus, JSOs were not preselected but represented a heterogeneous sample in terms of recidivism risk, offense characteristics, and ACEs. We aimed at approximating the dark figure of juvenile crime by not relying only on recidivism data from official registers but also including reoffending that had not been legally charged. Analyses considered different recidivism periods to examine short- and long-term risk prediction and were based on different methodological approaches in order to confirm our findings.

However, the present study is not without limitations. Because investigated case files had not been originally prepared for research purposes, information on some variables may have been limited, leading to a potential under-estimation of risk factors (e.g., concerning a JSO's perception of remorse or responsibility). Furthermore, despite considering both official and non-registered reoffending, we cannot rule out that some re-offenses may not have been identified. Although the inclusion of non-contact offenses and non-registered reoffending may have led to a more realistic representation of juvenile sexual offending and to an approximation toward the dark figure of crime, it also reduces the comparability to other studies that have only included contact offending and re-offenses of which a JSO has been found guilty according to different (internationally diverse) penal codes.

Several researchers have found that subscales of the J-SOAP II differed in their predictive validities concerning criminal recidivism (e.g., Aebi et al., 2011). However, as the

aim of the present study was to compare the overall scales of different instruments, the investigation of specific subscales was beyond the scope of the present study.

Due to staff shortage, recidivism (from case files) was rated by the same persons who had scored other variables of interest, which might have been a potential source of bias. As case files were not originally prepared for research purposes, some information was difficult to rate which lead to reduced inter-rater agreement. Because most case files had been created several years ago and thus responsible persons were not available for any re-checking, some ambiguities were not solvable. The present inter-rater agreement, however, supports the use of all variables of interest in the present study. Yet, it must be acknowledged that the present thresholds for inter-rater agreement (Fleiss, 1981; Landis & Koch, 1977) may be sufficient for research purposes, but somewhat higher thresholds have been required for applied/forensic settings (e.g., Edens & Boccaccini, 2017; Nunnally & Bernstein, 1994).

Finally, predictive validity was defined in terms of discrimination whereas calibration was not investigated. Although often found in studies in this area, this approach has been considered as fragmentary in evaluating the predictive accuracy of risk assessment scores (Singh, 2013). However, no reliable measures for calibration are available to date for instruments without specific cut-off values like those used in the present study (Singh, 2013). Future research may contribute to the field by examining whether the cut-off values of the VRAG-R for adult offenders are applicable to juvenile offenders too, and whether similar cut-off values can be derived for other risk assessment tools like the J-SOAP II. Furthermore, previous research has pointed to somewhat different accuracies of risk assessment scores with regard to theoretically defined JSO subsamples (e.g., JSOs with and without histories of general delinquent behaviors; Rajlic & Gretton, 2010). Taking into consideration the growing interest in empirically-derived subtypes of delinquent youth (e.g., Barra et al., 2017; Mulder, Vermunt, Brand, Bullens, & Marle, 2012), research on the performance of risk assessment

tools in distinct JSO subtypes appears promising in order to clarify which instruments are of specific usefulness for particular JSOs (e.g., JSOs with non-contact vs. contact offending).

Conclusions

The present study supports the use of the J-SOAP II, the ERASOR, and the VRAG-R in risk prediction of JSOs. Furthermore, the present findings show the importance of offense severity and the number of ACEs in risk assessment of youth. In particular the findings of the ACE score and their interactions with risk scores may challenge the validity of risk assessment instruments in highly burdened JSO. The heterogeneity of ACEs, coupled with different combinations of static and dynamic risk factors are important considerations for future research. Risk assessment and related consequences for an individual JSO must not be – at this juncture – based on respective instruments alone but must integrate a comprehensive consideration of a JSO's severity of offenses and history of psychosocial adversities in order to offer JSOs interventions that match their recidivism risks and individual needs (see, e.g., the risk-need-responsivity model; Andrews & Bonta, 2010).

References

Aebi, M., Landolt, M. A., Mueller-Pfeiffer, C., Schnyder, U., Maier, T., & Mohler-Kuo, M. (2015). Testing the "sexually abused-abuser hypothesis" in adolescents: A population-based study. *Archives of Sexual Behavior, 44*(8), 2189-2199. doi:10.1007/s10508-014-0440-x

Aebi, M., Plattner, B., Steinhausen, H. C., & Bessler, C. (2011). Predicting sexual and nonsexual recidivism in a consecutive sample of juveniles convicted of sexual offences. *Sexual Abuse: a Journal of Research and Treatment, 23*(4), 456-473. doi:10.1177/1079063210384634

Andrews, D. A., & Bonta, J. (2010). *The psychology of criminal conduct.* New Providence, NJ: LexisNexis.

Aylwin, A. S., Clelland, S. R., Kirkby, L., Reddon, J. R., Studer, L. H., & Johnston, J. (2000). Sexual offense severity and victim gender preference: A comparison of adolescent and adult sex offenders. *International Journal of Law and Psychiatry, 23*(2), 113-124. doi:10.1016/S0160-2527(99)00038-2

Baglivio, M. T., Wolff, K. T., Piquero, A. R., & Epps, N. (2015). The relationship between adverse childhood experiences (ACE) and juvenile offending trajectories in a juvenile offender sample. *Journal of Criminal Justice, 43*(3), 229-241.

Barra, S., Bessler, C., Landolt, M. A., & Aebi, M. (2017). Patterns of adverse childhood experiences in juveniles who sexually offended. *Sexual Abuse: a Journal of Research and Treatment*. Advance online publication. doi:10.1177/1079063217697135

Boer, D. P. (1997). *Manual for the Sexual Violence Risk-20: Professional guidelines for assessing risk of sexual violence*. Vancouver, Canada: British Columbia Institute Against Family Violence.

Brown, J., & Singh, J. P. (2014). Forensic risk assessment: A beginner's guide. *Archives of Forensic Psychology, 1*(1), 49-59.

Burton, D. L. (2003). Male adolescents: Sexual victimization and subsequent sexual abuse. *Child and Adolescent Social Work Journal, 20*(4), 277-296. doi:10.1023/A:1024556909087

Caldwell, M. F. (2016). Quantifying the decline in juvenile sexual recidivism rates. *Psychology, Public Policy, and Law, 22*(4), 414-426. doi:10.1037/law0000094

Caldwell, M. F., Ziemke, M. H., & Vitacco, M. J. (2008). An examination of the sex offender registration and notification act as applied to juveniles: Evaluating the ability to predict sexual recidivism. *Psychology, Public Policy, and Law, 14*(2), 89. doi: 10.1037/a0013241

Carpentier, J., & Proulx, J. (2011). Correlates of recidivism among adolescents who have

sexually offended. *Sexual Abuse: a Journal of Research and Treatment, 23*(4), 434-455. doi:10.1177/1079063211409950

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences*. Hillside, NJ: Lawrence Earlbaum Associates.

Edens, J. F., & Boccaccini, M. T. (2017). Taking forensic mental health assessment "out of the lab" and into "the real world": Introduction to the special issue on the field utility of forensic assessment instruments and procedures. *Psychological Assessment, 29*(6), 599-610. doi:10.1037/pas0000475

Fanniff, A. M., & Letourneau, E. J. (2012). Another piece of the puzzle: Psychometric properties of the J-SOAP-II. *Sexual Abuse: a Journal of Research and Treatment, 24*(4), 378-408. doi:10.1177/1079063211431842

Federal Bureau of Investigation. (2016). *2015 crime in the united states - table 32.* [Data file]. Retrieved from: https://ucr.fbi.gov/crime-in-the-u.s/2015/crime-in-the-u.s.-2015/tables/table-32

Fleiss, J. L. (1981). *Statistical methods for raters and proportions*. New York, NY: John Wiley & Sons, Inc.

Forth, A. E., Kosson, D. S., & Hare, R. D. (2003). *The hare psychopathy checklist: Youth version*. North Tonawanda, NY: Multi-Health Systems.

Fox, B. H., Perez, N., Cass, E., Baglivio, M. T., & Epps, N. (2015). Trauma changes everything: Examining the relationship between adverse childhood experiences and serious, violent and chronic juvenile offenders. *Child Abuse & Neglect, 46*, 163-173. doi:10.1016/j.chiabu.2015.01.011

Glover, A. J. J., Churcher, F. P., Gray, A. L., Mills, J. F., & Nicholson, D. E. (2017). A cross-

validation of the violence risk appraisal guide-revised (VRAG-R) within a correctional sample. *Law and Human Behavior*. Advance online publication. doi:10.1037/lhb0000257

Hair, J. F. J., Anderson, R. E., Tatham, R. L., & Black, W. C. (1995). *Multivariate data analysis (3rd ed.)*. New York, NY: Macmillan.

Hanley, J. A., & McNeil, B. J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology, 148*(3), 839-843. doi:10.1148/radiology.148.3.6878708

Hanson, R. K., & Morton-Bourgon, K. E. (2009). The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis of 118 prediction studies. *Psychological Assessment, 21*(1), 1-21. doi:10.1037/a0014421

Hare, R. D. (2003). *The psychopathy checklist–revised.* Toronto, Canada: Multi-Health Systems.

Harris, G. T., Rice, M. E., & Quinsey, V. L. (1993). Violent recidivism of mentally disordered offenders. *Criminal Justice and Behavior, 20*(4), 315-335. doi:10.1177/0093854893020004001

Harris, G. T., Rice, M. E., Quinsey, V. L., & Cormier, C. A. (2015). *Violent offenders: Appraising and managing risk (3rd ed.)*. Washington, DC: American Psychological Association.

Hecker, J. E. (2014). Baby with the bath water: Response to Fanniff and Letourneau. *Sexual Abuse: a Journal of Research and Treatment, 26*(5), 395-400. doi:10.1177/1079063214525644

Hempel, I., Buck, N., Cima, M., & van Marle, H. (2013). Review of risk assessment

instruments for juvenile sex offenders: What is next? *International Journal of Offender Therapy and Comparative Criminology, 57*(2), 208-228. doi:10.1177/0306624x11428315

Hogan, N. R., & Olver, M. E. (2016). Assessing risk for aggression in forensic psychiatric inpatients: An examination of five measures. *Law and Human Behavior, 40*(3), 233-243. doi:10.1037/lhb0000179

Howard, P. D. (2016). The effect of sample heterogeneity and risk categorization on area under the curve predictive validity metrics. *Criminal Justice and Behavior, 44*(1), 103-120. doi:10.1177/0093854816678899

Isele, D., Teicher, M. H., Ruf-Leuschner, M., Elbert, T., Kolassa, I. T., Schury, K., & Schauer, M. (2014). KERF - Ein Instrument zur umfassenden Ermittlung belastender Kindheitserfahrungen. Erstellung und psychometrische Beurteilung der deutschsprachigen MACE (Maltreatment and Abuse Chronology of Exposure) Scale [KERF–an instrument for measuring adverse childhood experiences: Construction and psychometric evaluation of the german MACE (Maltreatment and Abuse Chronology of Exposure) scale]. *Zeitschrift für Klinische Psychologie und Psychotherapie, 43*(2), 121-130. doi:10.1026/1616-3443/a000257

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159-174. doi:10.2307/2529310

Leroux, E. J., Pullman, L. E., Motayne, G., & Seto, M. C. (2016). Victim age and the generalist versus specialist distinction in adolescent sexual offending. *Sexual Abuse: a Journal of Research and Treatment, 28*(2), 79-95. doi:10.1177/1079063214535814

Mallie, A., Viljoen, J. L., Mordell, S., Spice, A., & Roesch, R. (2011). Childhood abuse and adolescent sexual re-offending: A meta-analysis. *Child & Youth Care Forum, 40*(5), 401-417. doi:10.1007/s10566-010-9136-0

Martinez, R., Rosenfeld, B., Cruise, K., & Martin, J. (2015). Predictive validity of the J-SOAP-II: Does accuracy differ across settings? *International Journal of Forensic Mental Health, 14*(1), 56-65. doi:10.1080/14999013.2015.1019683

Miccio-Fonseca, L. C. (2016). What is the recidivism rate of adolescent sex offenders? *Perspectives: California Coalition on Sexual Offending (CCOSO) Quarterly Newsletter, 1*, 4-7.

Miccio-Fonseca, L. C., & Rasmussen, L. A. (2011). A concise review on validated risk assessment tools for sexually abusive youth. *Sexual Offender Treatment, 6*(2), 1-13.

Mohler-Kuo, M., Landolt, M. A., Maier, T., Meidert, U., Schönbucher, V., & Schnyder, U. (2014). Child sexual abuse revisited: A population-based cross-sectional study among Swiss adolescents. *Journal of Adolescent Health, 54*(3), 304-311.e1. doi:10.1016/j.jadohealth.2013.08.020

Mossman, D. (1994). Assessing predictions of violence: Being accurate about accuracy. *Journal of Consulting and Clinical Psychology, 62*(4), 783-792. doi:10.1037/0022-006X.62.4.783

Mulder, E., Vermunt, J., Brand, E., Bullens, R., & Marle, H. (2012). Recidivism in subgroups of serious juvenile offenders: Different profiles, different risks? *Criminal Behaviour and Mental Health, 22*(2), 122-135. doi:10.1002/cbm.1819

Murphy, W. D., Page, I. J., & Hoberman, H. M. (2016). Adolescents who have engaged in sexually abusive behavior: An overview. In A. Phenix & H. M. Hoberman (Eds.), *Sexual offending: Predisposing antecedents, assessments and management* (pp. 185-212). New York, NY: Springer New York.

Nedopil, N., Grassl, P., & Mende, W. (1986). [The forensic psychiatric documentation system (FPDS). Development and initial application in penal expertise]. *Acta Psychiatrica Belgica, 87*(1), 98-113.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory (3rd ed.)*. New York, NY: McGraw-Hill.

Parks, G. A., & Bard, D. E. (2006). Risk factors for adolescent sex offender recidivism: Evaluation of predictive factors and comparison of three groups based upon victim type. *Sexual Abuse: a Journal of Research and Treatment, 18*(4), 319-342. doi:10.1177/107906320601800402

Prentky, R. A., Harris, B., Frizzell, K., & Righthand, S. (2000). An actuarial procedure for assessing risk with juvenile sex offenders. *Sexual Abuse: a Journal of Research and Treatment, 12*(2), 71-93. doi:10.1023/A:1009568006487

Prentky, R. A., Li, N. C., Righthand, S., Schuler, A., Cavanaugh, D., & Lee, A. F. (2010). Assessing risk of sexually abusive behavior among youth in a child welfare sample. *Behavioral Sciences & the Law, 28*(1), 24-45. doi:10.1002/bsl.920

Prentky, R. A., & Righthand, S. (2003). *Juvenile sex offender assessment protocol-II (J-SOAP-II) manual*. Washington, DC: US Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention.

Quenzer, C., & Dahle, K.-P. (2010). Standardisierte Prognoseinstrumente für junge Sexualstraftäter [standardized risk instruments for juvenile sexual offenders]. In P. Briken, A. Spehr, G. Romer, & W. Berner (Eds.), *Sexuell grenzverletzende Kinder und Jugendliche* (pp. 329-341). Lengerich, Germany: Pabst Science.

Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (Eds.). (1998). *Violent offenders: Appraising and managing risk (1st ed.)*. Washington, DC: American Psychological Association.

Rajlic, G., & Gretton, H. M. (2010). An examination of two sexual recidivism risk measures in adolescent offenders. *Criminal Justice and Behavior, 37*(10), 1066-1085. doi:10.1177/0093854810376354

Ralston, C. A., & Epperson, D. L. (2013). Predictive validity of adult risk assessment tools with juveniles who offended sexually. *Psychological Assessment, 25*(3), 905-916. doi:10.1037/a0032683

Rasmussen, L. A. (2013). Young people who sexually abuse: A historical perspective and future directions. *Journal of Child Sexual Abuse, 22*(1), 119-141. doi:10.1080/10538712.2013.744646

Rettenberger, M., & Eher, R. (2007). Predicting reoffense in sexual offender subtypes: A prospective validation study of the German version of the Sexual Offender Risk Appraisal Guide (SORAG). *Sexual Offender Treatment, 2*(2), 1-12.

Rettenberger, M., Gregório Hertz, P., & Eher, R. (2017). *Die deutsche Version des Violence Risk Appraisal Guide-Revised (VRAG-R) [The German version of the Violence Risk Appraisal Guide-Revised (VRAG-R)]*. Wiesbaden, Germany: Kriminologische Zentralstelle.

Rettenberger, M., Klein, V., Martin, R., & Briken, P. (2014). Die Einschätzung des Rückfallrisikos bei sexuell auffälligen Kindern und Jugendlichen anhand standardisierter Prognoseinstrumente [The accuracy of standardized risk assessment instruments for the prediction of recidivism in accused juveniles who sexually offended]. *Kindheit und Entwicklung, 23*(4), 210-219. doi: 10.1026/0942-5403/A000146

Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: Roc area, Cohen's d, and r. *Law and Human Behavior, 29*(5), 615-620. doi:10.1007/s10979-005-6832-7

Rice, M. E., Harris, G. T., & Lang, C. (2013). Validation of and revision to the VRAG and SORAG: The violence risk appraisal guide-revised (VRAG-R). *Psychological Assessment, 25*(3), 951-965. doi:10.1037/a0032878

Righthand, S., Prentky, R., Knight, R., Carpenter, E., Hecker, J., & Nangle, D. (2005). Factor structure and validation of the juvenile sex offender assessment protocol (J-SOAP). *Sexual Abuse: a Journal of Research and Treatment, 17*(1), 13-30. doi:10.1007/s11194-005-1207-7

Ryan, E. P. (2016). Juvenile sex offenders. *Child and Adolescent Psychiatric Clinics of North America, 25*(1), 81-97. doi:10.1016/j.chc.2015.08.010

Schlank, A., Matheny, S., & Schilling, J. (2016). Overview of assessment of sexual offenders. In A. Phenix & H. M. Hoberman (Eds.), *Sexual offending: Predisposing antecedents, assessments and management* (pp. 247-264). New York, NY: Springer New York.

Schmelzle, M. (2003). *Skala zur Einschätzung des Rückfallrisikos bei Jugendlichen Sexualstraftätern - Deutsche Übersetzung Version 0.0/2003 des ERASOR: Estimate of Risk of Adolescent Sexual Offense Recidivism Version 2.0 2001 [Skala zur Einschätzung des Rückfallrisikos bei Jugendlichen Sexualstraftätern – German translation version 0.0/2003 of the ERASOR: Estimate of Risk of Adolescent Sexual Offense Recidivism Version 2.0 2001]*. Retrieved from https://www.forio.ch/service/erasor

Schmelzle, M. (2004). *Juvenile sex offender assessment protocol II (J-SOAP-II), manual*. Retrieved from https://www.forio.ch/service/j-soap

Seto, M. C., & Lalumière, M. L. (2010). What is so special about male adolescent sexual offending? A review and test of explanations through meta-analysis. *Psychological Bulletin, 136*(4), 526-575. doi:10.1037/a0019700

Singh, J. P. (2013). Predictive validity performance indicators in violence risk assessment: A methodological primer. *Behavioral Sciences & the Law, 31*(1), 8-22. doi:10.1002/bsl.2052

Singh, J. P., Grann, M., & Fazel, S. (2011). A comparative study of violence risk assessment tools: A systematic review and metaregression analysis of 68 studies involving 25,980 participants. *Clinical Psychology Review, 31*(3), 499-513. doi:10.1016/j.cpr.2010.11.009

Swiss Federal Statistical Office. (2016). *Polizeilich registrierte Beschuldigte gemaess Strafgesetzbuch nach Kanton, Aufenthaltssgruppe, Altersklasse und Geschlecht [Alleged criminals registered by the police according to canton, resident permit status, age, and sex]*. Retrieved from: https://www.bfs.admin.ch/bfs/de/home/statistiken/kriminalitaet-strafrecht/polizei/beschuldigte.assetdetail.333820.html

Teicher, M. H., & Parigger, A. (2015). The 'Maltreatment and Abuse Chronology of Exposure' (MACE) scale for the retrospective assessment of abuse and neglect during development. *PLoS One, 10*(2), e0117423. doi:10.1371/journal.pone.0117423

Thornberry, T. P., & Krohn, M. D. (2000). The self-report method for measuring delinquency and crime. *Criminal justice, 4*(1), 33-83.

van der Put, C. E., & De Ruiter, C. (2016). Child maltreatment victimization by type in relation to criminal recidivism in juvenile offenders. *BMC Psychiatry, 16*, 24. doi:10.1186/s12888-016-0731-y

van der Put, C. E., van Vugt, E. S., Stams, G. J. J. M., Deković, M., & van der Laan, P. H. (2013). Short-term general recidivism risk of juvenile sex offenders: Validation of the Washington State juvenile court prescreen assessment. *International Journal of Offender Therapy and Comparative Criminology, 57*(11), 1374-1392. doi:10.1177/0306624x12457775

Viljoen, J. L., Elkovitch, N., Scalora, M. J., & Ullman, D. (2009). Assessment of reoffense

risk in adolescents who have committed sexual offenses: Predictive validity of the ERASOR, PCL:YV, YLS/CMI, and STATIC-99. *Criminal Justice and Behavior, 36*(10), 981-1000. doi:10.1177/0093854809340991

Viljoen, J. L., Mordell, S., & Beneteau, J. L. (2012). Prediction of adolescent sexual reoffending: A meta-analysis of the J-SOAP-II, ERASOR, J-SORRAT-II, and STATIC-99. *Law and Human Behavior, 36*(5), 423-438. doi:10.1037/h0093938

Viljoen, J. L., Scalora, M., Cuadra, L., Bader, S., Chávez, V., Ullman, D., & Lawrence, L. (2008). Assessing risk for violence in adolescents who have sexually offended. *Criminal Justice and Behavior, 35*(1), 5-23. doi:10.1177/0093854807307521

Wijetunga, C., Martinez, R., Rosenfeld, B., & Cruise, K. (2016). The influence of age and sexual drive on the predictive validity of the Juvenile Sex Offender Assessment Protocol–Revised. *International Journal of Offender Therapy and Comparative Criminology*. Advance online publication. doi:10.1177/0306624x16650681

Worling, J. R. (2004). The Estimate of Risk of Adolescent Sexual Offense Recidivism (ERASOR): Preliminary psychometric data. *Sexual Abuse: a Journal of Research and Treatment, 16*(3), 235-254. doi:10.1177/107906320401600305

Worling, J. R., Bookalam, D., & Litteljohn, A. (2012). Prospective validity of the estimate of risk of adolescent sexual offense recidivism (ERASOR). *Sexual Abuse: a Journal of Research and Treatment, 24*(3), 203-223. doi:10.1177/1079063211407080

Worling, J. R., & Curwen, T. (2001). *Estimate of Risk of Adolescent Sexual Offense Recidivism, Version 2.0*. Toronto, Canada: Ontario Ministry of Community and Social Services.

Worling, J. R., & Langton, C. M. (2015). A prospective investigation of factors that predict

desistance from recidivism for adolescents who have sexually offended. *Sexual Abuse: a Journal of Research and Treatment, 27*(1), 127-142. doi:10.1177/1079063214549260

Footnotes

[1] We did not include illegal pornography use in our definition of sexual (re)offending, because it does not necessarily involve the engagement of another person who is directly affected by the JSO's behavior.

[2] Before the assessment of inter-rater reliability, raters took part in training sessions in which several cases were first independently rated and then discussed afterwards. Based on these discussions, descriptions of the variables in the codebook were adapted in order to solve potential disagreement.

[3] We added a supplemental table with AUC results only for officially registered reoffending. Although most values are slightly smaller than those based on registered and non-registered reoffending, results are rather comparable.

[4] Rates of registered nonsexual-violent and general recidivism were sometimes higher than those from case file information for re-offenses committed up to 3 years and/or above age 18. This finding may be influenced by the fact that case files may not have covered long-term criminal conduct, because they contained information with greater temporal proximity to the initial offense.

Table 1

Pearson's/Spearman-Rank correlations among risk assessment instruments, offense severity scale, and cumulated adverse childhood experiences.

	J-SOAP II	ERASOR	VRAG-R	Offense Severity	ACE Score
J-SOAP II	1	.818***/.809***	.700***/.670***	.435***/.354***	.656***/.635***
ERASOR		1	.655***/.633***	.304***/.317***	.556***/.587***
VRAG-R			1	.231***/.232***	.508***/.508***
Offense Severity				1	.275***/.307***
ACE Score					1

Note. Total $N = 597$. ACE = Adverse childhood experiences.

*** $p \leq .001$.

Table 2

Rates of sexual, nonsexual-violent, and general recidivism based on different time periods and sources of information

	Recidivism rates: *n*(%)								
	Within 0.5 years (*n* = 597)			Within 3 years (*n* = 514)			Above age 18 (*n* = 550)		
	Total	Case files	Registered	Total	Case files	Registered	Total	Case files	Registered
Sexual	21 (3.5)	19 (3.2)	6 (1.0)	38 (7.4)	34 (6.6)	16 (3.1)	17 (3.1)	11 (2.0)	9 (1.6)
Non-Sexual violent	48 (8.0)	33 (5.5)	25 (4.2)	95 (18.5)	67 (13.0)	63 (12.3)	48 (8.7)	14 (2.5)	42 (7.6)
General	129 (21.6)	95 (15.9)	68 (11.4)	223 (43.4)	145 (28.2)	177 (34.4)	179 (32.5)	71 (12.9)	160 (29.1)

Note. Total *N* = 597.

Table 3

AUC values from ROC analyses including sexual, nonsexual-violent, and general recidivism for the three risk assessment instruments at different recidivism periods

Time period	Recidivism category	J-SOAP II			ERASOR			VRAG-R		
		AUC	95% CI		*AUC*	95% CI		*AUC*	95% CI	
			LL	*UL*		*LL*	*UL*		*LL*	*UL*
Within 0.5 years[a]	Sexual	.738***	.635	.841	.762***	.673	.851	.666*	.554	.777
	Nonsexual-violent	.750***	.680	.821	.692***	.613	.772	.726***	.641	.810
	General	.738***	.691	.785	.705***	.656	.753	.733***	.683	.782
Within 3 years[b]	Sexual	.740***	.662	.817	.780***	.707	.854	.694***	.599	.790
	Nonsexual-violent	.767***	.716	.818	.711***	.656	.766	.734***	.674	.794
	General	.713***	.669	.758	.680***	.633	.726	.718***	.673	.763
Above age 18[c]	Sexual	.701**	.571	.831	.703**	.572	.834	.688**	.528	.849
	Nonsexual-violent	.758***	.686	.829	.707***	.632	.782	.784***	.717	.852
	General	.632***	.583	.682	.635***	.585	.685	.660***	.611	.709

Note. *AUC* = area under the curve; CI = confidence interval; *LL* = lower limit; *UL* = upper limit.
[a]$n = 597$, [b]$n = 514$, [c]$n = 550$.
***$p \leq .001$, **$p \leq .01$, *$p \leq .05$.

Table 4

Univariate and multivariate Cox regressions concerning the prediction of sexual, nonsexual-violent, and general recidivism of the J-SOAP II

Indicators	Recidivism category								
	Sexual			Nonsexual-violent			General		
		95% CI			95% CI			95% CI	
	HR	*LL*	*UL*	*HR*	*LL*	*UL*	*HR*	*LL*	*UL*
Univariate Models									
J-SOAP II	1.97***	1.50	2.58	2.18***	1.84	2.60	1.73***	1.55	1.94
Multivariate Models									
Model 1									
J-SOAP II	2.07***	1.55	2.76	2.40***	1.99	2.89	1.79***	1.59	2.01
Severity score	0.86	0.62	1.19	0.76**	0.61	0.93	0.90	0.78	1.02
Model 2									
J-SOAP II	2.15***	1.61	2.86	2.41***	2.00	2.89	1.79***	1.59	2.02
Severity score	1.11	0.76	1.63	0.80	0.62	1.04	0.90	0.78	1.03
Interaction (J-SOAP II x Severity score)	0.70*	0.51	0.95	0.93	0.77	1.11	1.00	0.89	1.12
Model 3									
J-SOAP II	1.87***	1.29	2.71	1.97***	1.56	2.48	1.61***	1.39	1.87
ACE score	1.07	0.76	1.51	1.16	0.94	1.43	1.11	0.96	1.28
Model 4									
J-SOAP II	2.21***	1.51	3.24	2.07***	1.62	2.63	1.67***	1.44	1.95
ACE score	1.55*	1.05	2.30	1.30*	1.00	1.70	1.23*	1.04	1.44
Interaction (J-SOAP II x ACE score)	0.69*	0.52	0.92	0.90	0.77	1.06	0.89*	0.80	0.99
Model 5									
J-SOAP II	2.40***	1.61	3.58	2.27***	1.77	2.92	1.74***	1.49	2.04
Severity score	1.06	0.71	1.61	0.77^{+}	0.59	1.00	0.86*	0.74	1.00
ACE score	1.50*	1.00	2.24	1.32*	1.01	1.73	1.26**	1.06	1.48
Interaction (J-SOAP II x Severity score)	0.72	0.51	1.02	0.96	0.79	1.18	1.04	0.92	1.18
Interaction (J-SOAP II x ACE score)	0.71*	0.53	0.96	0.89	0.75	1.06	0.87*	0.77	0.98

Note. HR = hazard ratio, CI = confidence interval, *LL* = lower limit, *UL* = upper limit.
$^{+}p \leq .06$, *$p \leq .05$, **$p \leq .01$. ***$p \leq .001$.

Table 5

Univariate and multivariate Cox regressions concerning the prediction of sexual, nonsexual-violent, and general recidivism of the ERASOR

Indicators	Recidivism category								
	Sexual			Nonsexual-violent			General		
		95% CI			95% CI			95% CI	
	HR	*LL*	*UL*	*HR*	*LL*	*UL*	*HR*	*LL*	*UL*
Univariate Models									
ERASOR	2.65***	1.91	3.67	2.07***	1.71	2.50	1.70***	1.51	1.91
Multivariate Models									
Model 1									
ERASOR	2.74***	1.96	3.81	2.15***	1.77	2.62	1.73***	1.53	1.96
Severity score	0.86	0.63	1.18	0.85	0.69	1.04	0.93	0.82	1.06
Model 2									
ERASOR	2.71***	1.94	3.77	2.13***	1.75	2.59	1.73***	1.53	1.95
Severity score	1.04	0.67	1.63	0.90	0.71	1.15	0.94	0.82	1.08
Interaction (ERASOR x Severity score)	0.82	0.58	1.16	0.91	0.75	1.12	0.97	0.86	1.10
Model 3									
ERASOR	2.59***	1.79	3.76	1.73***	1.38	2.16	1.53***	1.33	1.76
ACE score	1.04	0.77	1.39	1.36***	1.13	1.63	1.21**	1.07	1.37
Model 4									
ERASOR	2.62***	1.81	3.78	1.77***	1.41	2.21	1.54***	1.34	1.77
ACE score	1.21	0.75	1.94	1.47**	1.16	1.86	1.28***	1.10	1.49
Interaction (ERASOR x ACE score)	0.88	0.64	1.22	0.92	0.78	1.09	0.93	0.83	1.04
Model 5									
ERASOR	2.64***	1.83	3.82	1.81***	1.44	2.28	1.57***	1.36	1.81
Severity score	1.01	0.64	1.61	0.85	0.66	1.10	0.91	0.79	1.05
ACE score	1.20	0.74	1.95	1.50***	1.18	1.91	1.30***	1.12	1.51
Interaction (ERASOR x Severity score)	0.84	0.59	1.19	0.92	0.74	1.14	0.99	0.86	1.12
Interaction (ERASOR x ACE score)	0.91	0.66	1.26	0.92	0.78	1.10	0.92	0.82	1.04

Note. HR = hazard ratio, CI = confidence interval, *LL* = lower limit, *UL* = upper limit.
*$p \leq .05$, **$p \leq .01$. ***$p \leq .001$.

Table 6

Univariate and multivariate Cox regressions concerning the prediction of sexual, nonsexual-violent, and general recidivism of the VRAG-R

Indicators	Recidivism category								
	Sexual			Nonsexual-violent			General		
		95% CI			95% CI			95% CI	
	HR	*LL*	*UL*	*HR*	*LL*	*UL*	*HR*	*LL*	*UL*
Univariate Models									
VRAG-R	1.64***	1.30	2.07	1.96***	1.69	2.27	1.65***	1.50	1.83
Multivariate Models									
Model 1									
VRAG-R	1.65***	1.29	2.11	2.06***	1.75	2.41	1.68***	1.51	1.87
Severity score	0.97	0.71	1.33	0.84	0.69	1.03	0.94	0.83	1.07
Model 2									
VRAG-R	1.75***	1.37	2.24	2.12***	1.80	2.49	1.74***	1.56	1.95
Severity score	1.14	0.81	1.61	0.92	0.73	1.17	0.98	0.86	1.12
Interaction (VRAG-R x Severity score)	0.79^{+}	0.62	1.01	0.90	0.78	1.04	0.92	0.83	1.01
Model 3									
VRAG-R	1.43*	1.08	1.89	1.71***	1.44	2.04	1.51***	1.35	1.70
ACE score	1.33^{+}	1.00	1.77	1.37***	1.15	1.63	1.23***	1.09	1.39
Model 4									
VRAG-R	1.70***	1.28	2.26	1.76***	1.45	2.13	1.57***	1.39	1.78
ACE score	1.64***	1.21	2.22	1.42***	1.16	1.75	1.29***	1.13	1.48
Interaction (VRAG-R x ACE score)	0.74*	0.59	0.94	0.95	0.83	1.10	0.92	0.84	1.01
Model 5									
VRAG-R	1.77***	1.32	2.38	1.87***	1.53	2.28	1.63***	1.43	1.86
Severity score	1.06	0.73	1.53	0.86	0.67	1.09	0.94	0.82	1.08
ACE score	1.61**	1.18	2.21	1.42**	1.14	1.77	1.29***	1.12	1.48
Interaction (VRAG-R x Severity score)	0.82	0.63	1.08	0.92	0.79	1.07	0.95	0.85	1.05
Interaction (VRAG-R x ACE score)	0.77*	0.61	0.97	0.98	0.85	1.14	0.93	0.84	1.03

Note. HR = hazard ratio, CI = confidence interval, *LL* = lower limit, *UL* = upper limit.
$^{+}p \leq .06$, *$p \leq .05$, **$p \leq .01$. ***$p \leq .001$.